

**SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO – MATEMATIČKI FAKULTET
BIOLOŠKI ODSJEK**

**PORIJEKLO GENETIČKOG KODA
THE ORIGIN OF THE GENETIC CODE
SEMINARSKI RAD**

Dora Zvjerković
Preddiplomski studij Molekularne biologije
Undergraduate Study of Molecular Biology
Mentor: izv. prof. dr. sc. Damjan Franjević

Zagreb, 2019.

1. UVOD	3
2. „GOTOVO” UNIVERZALNI GENETIČKI KOD	6
2.1 ZAŠTO JE KOD (GOTOVO) UNIVERZALAN?	9
3. RANE I KASNE AMINOKISELINE.....	11
4. VODEĆE TEORIJE O PORIJEKLU SGC.....	12
4.1. STEREOKEMIJSKA TEORIJA	12
4.2. KOEVOLUCIJSKA TEORIJA	14
4.3. TEORIJA MINIMIZACIJE GREŠKE.....	16
5. KOKOŠ ILI JAJE?	19
6. ZAKLJUČAK.....	21
7. SAŽETAK	22
8. SUMMARY	23
9. LITERATURA.....	24

1. UVOD

Genetički kod svojevrsan je rječnik po kojemu se sekvenca nukleotida u genima posrednim mehanizmom vjerno prevodi u sekvencu aminokiselina koja čini protein. Kod je strukturiran na način da svaki od 64 mogućih tripleta RNA (kodona) sastavljenih od kombinacija 4 nukleotida (adenin, uracil, gvanin, citozin) odgovara jednoj od 20 biološki aktivnih aminokiselina, uz dodatna 3 „stop” kodona koji terminiraju translaciju proteina (Crick, 1968.).

UUU [F] Phe	UCU [S] Ser	UAU [Y] Tyr	UGU [C] Cys
UUC [F] Phe	UCC [S] Ser	UAC [Y] Tyr	UGC [C] Cys
UUA [L] Leu	UCA [S] Ser	UAA [] Ter	UGA [] Ter
UUG [L] Leu	UCG [S] Ser	UAG [] Ter	UGG [W] Trp
CUU [L] Leu	CCU [P] Pro	CAU [H] His	CGU [R] Arg
CUC [L] Leu	CCC [P] Pro	CAC [H] His	CGC [R] Arg
CUA [L] Leu	CCA [P] Pro	CAA [Q] Gln	CGA [R] Arg
CUG [L] Leu	CCG [P] Pro	CAG [Q] Gln	CGG [R] Arg
AUU [I] Ile	ACU [T] Thr	AAU [N] Asn	AGU [S] Ser
AUC [I] Ile	ACC [T] Thr	AAC [N] Asn	AGC [S] Ser
AUA [I] Ile	ACA [T] Thr	AAA [K] Lys	AGA [R] Arg
AUG [M] Met	ACG [T] Thr	AAG [K] Lys	AGG [R] Arg
GUU [V] Val	GCU [A] Ala	GAU [D] Asp	GGU [G] Gly
GUC [V] Val	GCC [A] Ala	GAC [D] Asp	GGC [G] Gly
GUA [V] Val	GCA [A] Ala	GAA [E] Glu	GGA [G] Gly
GUG [V] Val	GCG [A] Ala	GAG [E] Glu	GGG [G] Gly

Slika 1. Standardni genetički kod (SGC). Sjenčanje odgovara PR (Polar Requirement) skali, eksperimentalnoj mjeri polarnosti, radi naglaska na grupiranje kemijski sličnih aminokiselina. Polarnije aminokiseline obojene su svjetlije sivom bojom. Crni pravokutnici označavaju terminacijske (stop) kodone. Preuzeto iz Koonin i Novozhilov, 2009.

Standardni genetički kod (SGC) gotovo je potpuno konzerviran među svim trima carstvima života, pa je kao takav jedna od najočuvanijih informacijskih konstanti u čitavom biološkom svijetu (Koonin i sur., 2009.). Time je pitanje porijekla genetičkog koda nedvojbeno usko vezano uz pitanje

porijekla i rane evolucije samoga života na Zemlji, a ako se proširi i na pitanje porijekla odnosa nukleinskih i aminokiselina, može se reći i da se potpuno preklapaju. Stoga je iznimno zanimljivo što nije došlo do značajnih pomaka od hipoteze prema pravoj znanstvenoj, empirijski provjerljivoj teoriji u području istraživanja porijekla genetičkog koda u periodu od 5 desetljeća otkada je kod „dešifriran”, pa do danas.

S obzirom da je kod gotovo univerzalan (s manjim i rijetkim varijacijama koje će u ovom radu biti izložene kasnije), može se zaključiti da je kao takav, ili približno takav, bio etabliran već u LUCA-i (*Last Universal Common Ancestor*), zajedno s translacijskom mašinerijom sastavljenom od tRNA (transfer RNA), aaRS (aminoacil-tRNA sintetaza) i ribosoma (Koonin i sur., 2009.). Kako su navedene komponente evolucijski vrlo konzervirane, a fosili organizama koji su prethodili LUCA-i do danas nisu pronađeni (Martin i sur., 2016.), njihovu evoluciju je jako teško istraživati klasičnim metodama evolucijske biologije. Unatoč tome, uočene pravilnosti u kodu inspirirale su, i još uvijek inspiriraju, postavljanje raznih hipoteza i domišljatih eksperimenata kojima se pokušava otključati odgovor na temeljno pitanje evolucijske biologije – kako je nastao život.

Već pri letimičnom pogledu na tablicu Standardnoga genetičkog koda (Slika 1.), jasno je da nije nasumičan. U pravilu, kodoni koji se razlikuju u samo jednom nukleotidu, kodiraju prividno fizikalno-kemijski slične aminokiseline; prividno, zato što ne postoji definitivan kriterij za određivanje fizikalno-kemijske sličnosti (često korišten kriterij je *Polar requirement scale*, mjera topljivosti u piridinu (Ardell i sur., 2002.)). Posljedica ovakve grupacije je snažnija otpornost na supstitucije i greške pri translaciji; zamjena neke aminokiseline sličnom imat će manje negativan utjecaj na strukturu i funkciju proteina od zamjene vrlo različitom aminokiselinom (Sneath, 1966.). Pravokutnici u tablici Standardnog genetičkog koda su u većini slučajeva grupirane u skupove od 2 ili 4 pravokutnika – treći nukleotid kodona je većinom sinoniman tj. neinformativan (tzv. *wobble* pravilo koje je pojašnjeno kasnije u tekstu). Također je uočeno (Taylor, 1989.) da neke od aminokiselina koje se nalaze na istom anaboličkom putu kodiraju kodoni s istim prvim nukleotidom. Postoji 10^{84} mogućih alternativnih kodova (načina da se 64 kodona podijeli na 20 aminokiselina i terminacijski kodon, s uvjetom da je svakoj aminokiselini pridružen barem jedan kodon i da postoji barem jedan terminacijski kodon) (https://en.wikipedia.org/wiki/Genetic_code), dakle, iznimno je nevjerojatno da su primijećene pravilnosti u tablici koda gola slučajnost. Nameće se pitanje: kakva je točno sprega evolucijskih sila, selekcijskih kriterija, kemijskih i fizikalnih zakona i ograničenja, kombiniranih s nizom prapovijesnih slučajnosti, odgovorna za ovakav kod pun pravilnosti, uređen kod, očito nenasumičan kod?

Četiri teorije proizašle su iz mnogih kao najvažnije.

Teorija minimizacije greške, koja se čini intuitivnim zaključkom strukture tablice Standardnoga genetičkog koda, tvrdi da je SGC nastao iz nasumičnog koda koji je evoluirao pod selekcijskim pritiskom smanjenja grešaka uzrokovanih supstitucijama nukleotidnih baza u kodonima. Pretraživanjem nekih od mogućih nasumičnih kodova, Freeland i Hurst (1998.) došli su do slavnoga zaključka da je Standardni genetički kod „jedan u milijun”, tj. da je tek jedan od milijun nasumičnih kodova otporniji na greške od standardnoga. Koliko god impresivno to zvučalo, lako je uvidjeti da „jedan u milijun”, u prostoru od 10^{84} mogućih kodova, znači da još uvijek postoji otprilike 10^{14} potencijalnih kodova koji su otporniji na greške od SGC.

Također valja odgovoriti i na pitanje zašto su kodoni raspodijeljeni baš tako kako jesu, zašto npr. glicin ima kodone oblika GGN (gdje N označava bilo koji nukleotid), a ne, primjerice, GCN? Današnji mehanizam translacije, univerzalno konzerviran u svim živućim staničnim organizmima, uključuje enzime aaRS, koji aminoaciliraju tRNA odgovarajućim aminokiselinama. Dakle, mehanizam je posredan – ni u jednom koraku ne dolazi do izravnog prepoznavanja kodona ili antikodona od strane aminokiseline (Ibbs i sur., 2010.).

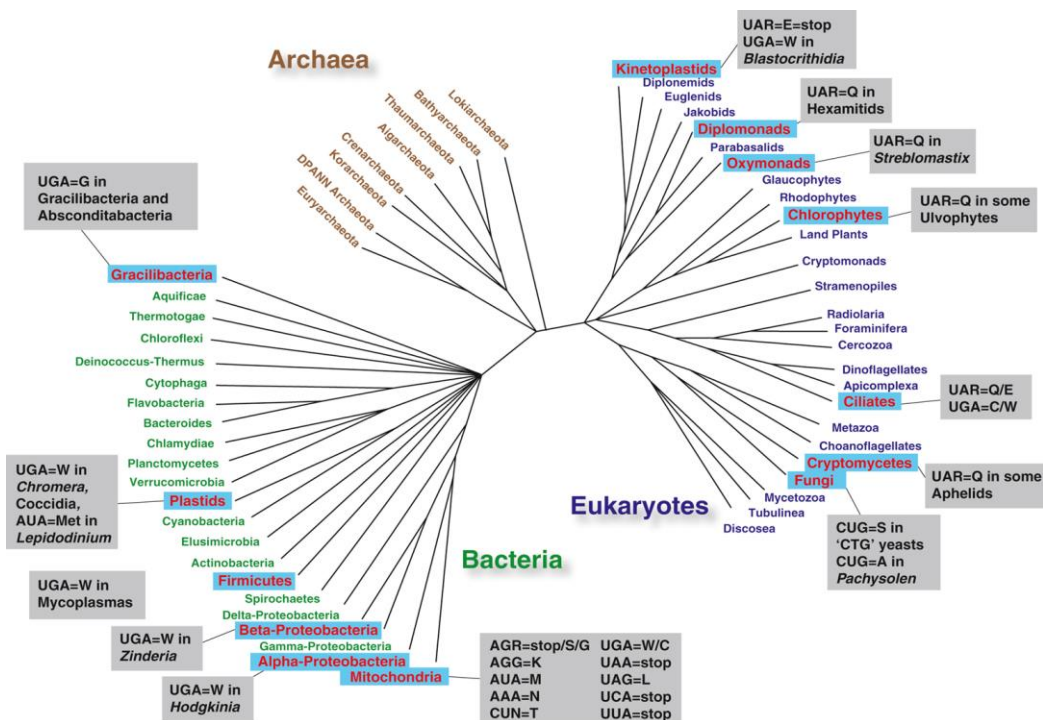
Stereokemijska teorija postavlja hipotezu da je u vremenu prije razvoja današnjeg posrednog mehanizma i razvoja specifičnih proteinskih enzima zapravo dolazilo do prepoznavanja između kodona i aminokiseline, te da današnji SGC odražava stereokemijske afinitete među kodonima i odgovarajućim aminokiselinama (Gamow, 1954.).

Pobornici **koevolucijske teorije** tvrde da je u najranijem periodu evolucije kod bio jednostavniji i raspolagao s manjim brojem ranih aminokiselina koje su mogle biti sintetizirane abiogenetski, a kod se razvijao zajedno s evolucijom metaboličkih puteva sinteze izvedenih aminokiselina. Prema tome bi struktura SGC odražavala metaboličke odnose između suvremenih 20 standardnih aminokiselina (Wong, 1975.).

Teorija „zamrznute slučajnosti” (eng. „*frozen accident*”) je jednostavna: ispočetka je istovremeno postojalo mnogo nasumičnih kodova od kojih je SGC bio jedan. Suvremeni kod je slučajnošću (eng. *random drift*) postao najbrojniji i istisnuo ostale kodove, a nakon što je ustanovljen u liniji organizama koja je dovela do LUCA-e, ikakve daljnje preinake koda imale su letalni učinak na fitnes proteina i prvih stanica; kod je ostao „zamrznut”. Teoriju zamrznute slučajnosti postavio je Crick (1968.) netom nakon otkrića SGC, a unatoč razvoju mnogih drugih, elegantnih i intuitivnih teorija, Crickova zamrznuta slučajnost čini se, s obzirom na manjak čvrstih empirijskih potvrda ostalih teorija, u nekom obliku jednako relevantnom i dan danas.

2. „GOTOVO” UNIVERZALNI GENETIČKI KOD

Kao što je prethodno spomenuto, genetički kod pokazuje određenu dozu nestandardnosti. Pronađeno je do sada više od 20 nestandardnih genetičkih kodova, uglavnom u organelima (kloroplasti, mitohondriji) i nekim parazitskim i endosimbiotskim organizmima s jako reduciranim genomima (Knight i sur., 2001.). Alternativni kodovi razlikuju se od standardnoga uglavnom u samo malom broju kodona. Postojanje varijacija je pomalo u kontradikciji s Crickovom teorijom zamrznute slučajnosti, jer demonstrira da je promjena koda moguća i da ne mora nužno biti letalna (s obzirom da se određene promjene pojavljuju više puta nezavisno, čini se da mogu biti i povoljne; pogledaj Sliku 2.). Unatoč tome, s obzirom da su varijacije minimalne (većinom se razlikuju od SGC u samo jednome kodonu), očito je da je evolucija koda ipak snažno ograničena i kod gotovo smrznut. Ove varijacije su izvedene; nestandardni kodovi nastali su evolucijom pod određenim selekcijskim pritiscima iz standardnoga, stoga nam ne mogu puno reći o periodu prije utemeljenja SGC. Dapače, prema jednom radu (Swire i sur., 2005.) evolucija genetičkog koda je podijeljena u dvije faze čije se karakteristike značajno razlikuju, a to su faza prije i faza poslije ustanovljenja SGC. Ipak nije naodmet proučiti prirodu spomenutih varijacija, jer nam one nešto razotkrivaju o načinima na koje se kod može mijenjati te o vrstama selekcije koje na nj mogu vršiti pritisak.

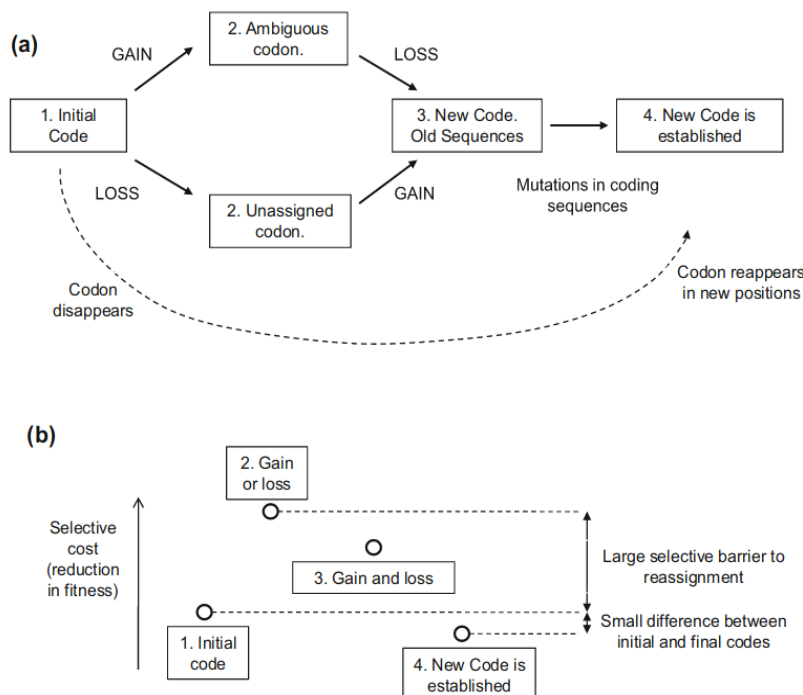


Current Biology

Slika 2. Varijacije u genetičkom kodu. Filogenetičko stablo života s naznačenim nekim od poznatih varijacija u genetičkom kodu u svim trima domenama života. Preuzeto iz Keeling, 2016.

Kvalitativna razlika među fazama prije i nakon ustanovljenja SGC može se objasniti točkom infleksije u funkciji koja opisuje povećanje fitnesa stanice ovisno o povećanju repertoara aminokiselina dostupnih za sintezu polipeptida. Ako uzmemo kao istinitu široko prihvaćenu pretpostavku da je kod prvotno raspolagao s manjim brojem aminokiselina, onda je prva faza evolucije SGC bila obilježena snažnom pozitivnom selekcijom na proširivanje raspona kiselina i, time, proširivanje mogućih funkcija proteina, dok je druga faza obilježena negativnim učinkom na fitnes koji sa sobom nosi prenamjena kodona. Ovakav način razmišljanja potencijalno objašnjava i zašto život raspolaže sa samo 20 (22) aminokiselina; moguće je da se ta brojka nalazi u ravnoteži između spomenutih pozitivnih i negativnih učinaka koje dodatak novih aminokiselina donosi. Također, drevne stanice vjerojatno su kodirale za znatno manji broj proteina od modernih stanica, čime bi promjene u kodu nosile manje opsežan negativni učinak na fitnes. Kako onda objasniti varijabilnost u kodovima koju primjećujemo u modernim organizmima?

Mehanizam većine promjena u kodonima svodi se na supstitucije u genima tRNA koje direktno utječu na dekodiranje (Sengupta i Higgs, 2005.) ili na modifikacije baza ili uređivanje tRNA zbog kojih ih prepoznaje drugačija aaRS od one koja pripada njezinom antikodonu prema tablici SGC. Poznato je i mnoštvo vrsta u čije proteine se ugrađuju dvije nestandardne aminokiseline: selenocistein i pirolizin. Moguće je da su promjene nastale postupno, i to jednim od dva moguća puta: preko dvoznačnog intermedijera i preko gubitka kodona (Slika 3.).



Slika 3. (a) Putevi prenamjene kodona preko dvoznačnog intermedijera ili prestanka korištenja kodona. (b) Shema gubitka fitnesa zbog promjene koda. Preuzeto iz Sengupta i Higgs, 2015.

Prvi je put onaj gdje dobitak novog značenja prethodi gubitku staroga na način da je u određenom periodu kodon dvoznačan, tj. dio vremena prevodi se u jednu, a dio u drugu aminokiselinu. Ovakav kod smanjuje fitness zato što će se barem dio vremena kodon prevoditi u 'krivu' aminokiselinu, tj. aminokiselinu koja smanjuje efikasnost proteina. Drugi je onaj gdje kodon prvo potpuno gubi značenje, ne prevodi se ni u jednu aminokiselinu ili se u nekom periodu uopće ne javlja u genomu, što također smanjuje fitness zato što stanica nema na raspolaganju tRNA koja odgovara kodonu ukoliko se taj nađe negdje u sekvenci. Oba puta zahtijevaju i opsežne mutacije u genomu kako bi se kodon s novo dobivenim značenjem počeo pojavljivati na mjestima gdje je povoljno da se prevodi u dotičnu aminokiselinu. Slika 3b shematski prikazuje selektivnu barijeru u putevima prenamjene kodona i razlog zašto su alternativni kodovi toliko rijetki.

Prema radu J. Swirea (2005.) koji je istraživao prirodu prenamjene kodona u 24 različita, nezavisno evoluirana mitohondrijska alternativna koda, slučajni drift može objasniti samo slučajeve gdje je neki od stop kodona zamijenjen kodonom za aminokiselinu, ali ne može objasniti slučajeve gdje je jedna aminokiselina zamijenjena drugom (Swire, 2005.). Naime, u većini potonjih slučajeva nije bila riječ o rijetko korištenom kodonu (kao što su svi stop kodoni), što je uvjet za to da drift bude vjerojatan, već su neki od češće korištenih kodona prenamijenjeni. Ovaj rad pruža drugačije objašnjenje: kodoni su prenamijenjeni prema metaboličkoj 'cijeni' aminokiselina. Takav selekcijski pritisak bio bi zanemariv u odnosu na drift i mutacijsku frekvenciju kad bi se radilo o evoluciji samo jedne aminokiseline u sekvenciji, no promjenom afiniteta tRNA prema jeftinijoj aminokiselini, učinci pozitivne selekcije aditivno se pospješuju u cijelom genomu gdje god se javlja taj kodon, te prevladavaju drift i mutaciju etablirajući novi genetički kod.

Analogno razmišljanje moglo bi se primijeniti i na ostale organizme s alternativnim genetičkim kodom, no ovo područje zahtjeva dodatno istraživanje da bismo došli do ikakvih definitivnih zaključaka. Ono što možemo primijetiti je da je kod fleksibilniji no što se možda intuitivno čini i da na njega može utjecati čitav niz različitih i međusobno neovisnih selekcijskih pritisaka, pa stoga jednostavna, izravna i intuitivna objašnjenja nisu uvijek primjenjiva u raspravama o evoluciji genetičkog koda.

2.1 ZAŠTO JE KOD (GOTOVO) UNIVERZALAN?

Zbog primjera alternativnih genetičkih kodova znamo da je SGC fleksibilan, može evoluirati i uistinu varira, no njega se svejedno često naziva univerzalnim zato što su sve (do sada pronađene) varijacije minimalne i generalna struktura koda je uvijek konzervirana. Zašto je SGC univerzalan? Scenarij zamrznute slučajnosti bavi se samo vremenom nakon što je kod već etabliran, no vrlo vjerojatno je prije utemeljenja SGC istovremeno postojalo mnogo drugih različitih kodova u kompeticiji. Zašto se samo SGC „zamrznuo” te ne nigdje u prirodi ne nalazimo potomke drugih prvotnih kodova?

Jedna ideja, izložena u radu Vestigiana iz 2006., je da je u samim počecima života horizontalni transfer gena (eng. *horizontal gene transfer*, nadalje HGT) bio esencijalni alat rane evolucije i da bez njega ne bi nikada moglo doći do staničnoga života. Prema ovoj hipotezi genetički kod bio je, osim što je spajao triplete nukleotida s pripadajućim aminokiselinama, i svojevrsni protokol po kojemu su zajednice ranih organizama mogle dijeliti inovacije – jasno je da bi HGT bio iznimno otežan, ako ne i onemogućen, u nedostatku zajedničkoga koda. Vestigian dijeli povijest života u 3 epohe.

U prvoj je život bio vrlo otporan na nepreciznosti. Aminokiseline nisu odgovarale specifičnim kodonima, već su grupe kodona odgovarale grupama aminokiselina (prema nekim, za ovu teoriju manje važnim, kemijskim ili fizikalnim pravilima) pa su tada polipeptidi bili stohastički sintetizirani. Komunalno dijeljenje inovacija, gena, i kompleksnih elemenata predstanične strukture bilo je široko rasprostranjeno, a iz takve dinamike proizašao je nužno i zajednički protokol, jezik kojim su te rane populacije organizama mogle međusobno komunicirati i evoluirati kolektivno – genetički kod. U ovoj fazi HGT bio je daleko opsežnije prirode no što ga danas shvaćamo i porijeklo se nije moglo pratiti vertikalno po paradigmi predak-potomak. Stoga ova teorija kolektivne evolucije ne priznaje koncept jedinstvenoga zajedničkog pretka.

U drugoj epohi, još uvijek široko rasprostranjen HGT potpomognut zajedničkim protokolom transfera (kodom) omogućio je brzi razvoj kompleksnosti. Kako kompleksnost zahtjeva nižu toleranciju na nepreciznosti i greške, kraj ove faze označava smanjenje ili prestanak opisane komunalne dinamike i konačno prijelaz na treću epohu obilježenu dinamikom vertikalnog prijenosa gena, prijelaz nazvan Darwinističkom tranzicijom. Dok je prvu i drugu epohu definirao HGT kao najvažniji mehanizam evolucijske inovacije, treću epohu, u kojoj se nalazimo i danas, definiraju mehanizmi Darwinističke evolucije i populacijske genetike; ovdje je stohastička priroda rane sinteze peptida potpuno iščezla, zamijenjena kompleksnom translacijskom mašinerijom, mehanizmima popravka grešaka u transkripciji i translaciji, obrane od strane DNA, te raznovrsnim mehanizmima rekombinacije i spolnog razmnožavanja

Pretpostavke dane u radu podržane su računalnim modeliranjem, ali i teoretski, eksperimentalno i računalno potvrđenom činjenicom da je HGT primarni faktor u adaptaciji i inovaciji i kod današnjih mikrobnih organizama (Levin i Bergstrom, 2000.). Klonalne populacije aseksualnih organizama u odsustvu bilo kakve rekombinacije (uključujući HGT) brzo podliježu letalnom efektu Mullerovog sita (*Muller's ratchet*) – procesu ireverzibilnog nakupljanja negativnih mutacija kroz generacije. Dakle, ovisnost aseksualnih populacija (kakve su bile, možemo s određenom sigurnošću reći, najranije populacije organizama) o HGT, i ovisnost HGT o zajedničkom kodu mogli su biti faktori koji su potaknuli univerzalnost koda. Međutim, opsežna podložnost HGT-u nosi sa sobom i rizik invazije parazitskih genetičkih elemenata, pa u određenim uvjetima važnost zaštite od tog rizika nadvlada nad dobrobiti HGT. Prethodno opisane varijacije u kodu javile su se ili u vrlo izoliranim populacijama gdje su bile fiksirane driftom, ili kod parazitskih populacija vrlo reduciranih genoma koje, sudeći po zakonima populacijske genetike, teže ka izumiranju, ili kod endosimbiotskih organizama i organela koji opstaju pod specifičnim selekcijskim pritiscima.

Dakle, prema trenutno postojećoj literaturi, dva su faktora pridonijela efektivnoj univerzalnosti SGC: negativni učinak na fitnes pri prenamjeni kodona, te važnost HGT za opstanak i evoluciju mikrobnih populacija.

3. RANE I KASNE AMINOKISELINE

Ni u jednom pokusu abiotske sinteze nije uspješno dobivena svaka od 20 proteinogenih aminokiselina. U klasičnom Miller-Uray pokusu 1952. godine dobiveno je njih 5, od čega su 3 proteinogene – Gly, Ala i Asp (Miller, 1953., Miller i Uray, 1959.). Međutim, nakon Mullerove smrti 2007. godine, ponovno je proučen sadržaj tikvica koje su zapečaćene još 1952. kada je pokus prvi puta izveden, i pronađeno je preko 20 različitih vrsta aminokiselina, od čega 10 proteinogenih (Bada, 2013.). Postoje razne nesuglasice oko sastava atmosfere i klimatoloških uvjeta na tadašnjoj Zemlji, te je trenutno generalni konsenzus da je atmosfera bila nešto manje reducirajuća i malo drugačijega sastava no što se mislilo kada je Miller dizajnirao svog pokus (Oparin, 1938., Kasting, 1993.). Svakako, mnoštvo pokusa prebiotske sinteze je provedeno u međuvremenu, a uz to su razne aminokiseline pronađene u palim meteoritima, te se konzistentno pronalazi 10 aminokiselina – kandidata za najstarije aminokiseline na Zemlji: Gly, Ala, Asp, Glu, Val, Ser, Ile, Leu, Pro i Thr (Trifonov, 2000.). Ne samo da su to i „najjeftinije” aminokiseline – s najmanjom Gibbsovom slobodnom energijom sinteze, već su i puno stabilnije od onih novijih, od kojih su mnoge ili termolabilne ili se lako degradiraju UV zračenjem (Kobayashi i sur., 1990.). Također, pokazano je filogenetičkom analizom paralognih proteina da se aminokiseline smatrane starijima kroz vrijeme sve rjeđe koriste u proteinima, dok se one smatrane novijima kroz vrijeme koriste sve češće (Brooks i sur., 2002.).

Sudeći po rečenome, čini se da su preostale aminokiseline morale biti sintetizirane pomoću nekakvih enzima, tek kada su one ranije već neko vrijeme postojale i polimerizirale, možda tek u nasumične polipeptide, ali vjerojatnije nakon što je neka vrst ranijeg koda već bila ustanovljena. Ovo će imati implikacije na neke od kasnije opisanih teorija o porijeklu genetičkoga koda.

4. VODEĆE TEORIJE O PORIJEKLU SGC

Uočene pravilnosti u strukturi SGC i njegova očita nenasumičnost kao da zahtijevaju objašnjenje koje ide dalje od jednostavne zamrznute slučajnosti. Tri najkoherentnije teorije o uzrocima tih pravilnosti, već spomenute koevolucijska, stereokemijska, i teorija minimizacije greške, bit će opisane u ovom poglavlju.

4.1. STEREOKEMIJSKA TEORIJA

Stereokemijska teorija bila je jedna od najranije postavljene (Gamow, 1954.), možda zbog njezine privlačne intuitivnosti; ona pruža direktan odgovor na pitanje odnosa aminokiselina i kodona, koji danas nije moguće izravno promatrati zbog posrednog mehanizma kojime se geni prevode u proteine. Međutim, svi rani pokušaji dizajniranja pokusa kojima bi se teorija provjerila bili su otežani činjenicom da ne postoji apsolutan kriterij za određivanje kemijskih značajki aminokiselina kojime bi se mogao odrediti njihov afinitet prema odgovarajućim kodonima ili antikodonima.

Veliki pomak došao je zahvaljujući razvitku tehnologije aptamera – kratkih sekvenci RNA specifičnog afiniteta koje se dobivaju iz smjese nasumičnih sekvenci selektivnim vezanjem za aminokiseline. U pokusima s aptamerima, pronađeno je barem za neke aminokiseline značajno povećan afinitet za aptamere koji sadrže njima pripadajuće kodone ili antikodone (Knight i Landweber, 1998.). U jednom od njih analizirano je 8 aminokiselina (Arg, Gln, His, Ile, Leu, Phe, Trp, Tyr), od kojih je čak 5 pokazivalo povećan afinitet za pripadajuće nukleotidne triplete (ili kodon ili antikodon); arginin je pokazao statistički najznačajniji afinitet (Knight, Landweber i Yarus, 2003.). Ovo je shvaćeno kao snažan pokazatelj valjanosti stereokemijske teorije; čak i slabi afiniteti mogli su kumulativno kroz godine snažno utjecati na povezivanje između aminokiselina i tripleta. Međutim, kao što je objašnjeno u trećem poglavlju ovoga rada, aminokiseline koje su pokazale najveći afinitet prema pripadajućim tripletima smatraju se novijim dodatcima aminokiselinskom repertoaru Zemlje, te su vjerojatno nastale tek nakon što je kod, barem u nekakvom jednostavnijem obliku, već bio ustanovljen i postojali su već nekakvi enzimi za njihovu sintezu. To možda nije neobično ako se u obzir uzme da te novije aminokiseline imaju veće i kompleksnije postrane lance koji imaju sposobnost specifičnije interakcije s oligonukleotidima. Još jedan zbunjujući rezultat ovih pokusa je to što su neke od istraženih aminokiselina pokazale afinitet za pripadajućim kodonom, a druge za pripadajućim antikodonom, a to neslaganje čini se kao da donosi samo poteškoće u pokušaju dokazivanja ovoga mehanizma kao bitnoga za ustanovljenje SGC.

Stereokemijskoj teoriji može se pristupiti i na drugi, manje direktan način. Analizom strukture rRNA i ribosomskih proteina, pokazano je da su u strukturama arhejskih i bakterijskih ribosoma aminokiseline značajno često prostorno blizu aminokiseline s pripadajućim tripletima – 11 aminokiselina pokazalo je značajnu prostornu bliskost svojim antikodonima, a 8 kodonima (Johnson

i Wang, 2010.). S obzirom da su ribosomi drevne strukture i među enzimima najkonzerviraniji u živom svijetu, ovaj pronalazak interpretiran je kao dokaz za određenu važnost stereokemijskog afiniteta za evoluciju koda. Interpretacija je uzela u obzir i to da su, opet, novije aminokiseline pokazale veću statističku značajnost prostorne bliskosti tripletima: prema ovoj interpretaciji evolucija koda tekla je u dvije faze. U prvoj je kod postojao u jednostavnijem obliku i imao je već ukomponirane aminokiseline pronađene u okolini organizama, koje nisu pridružene svojim tripletima po ikakvom stereokemijskom afinitetu već prema nekakvom drugom pravilu. U drugoj, nove sintetizirane aminokiseline polako su se ugrađivale u kod prema stereokemijskom afinitetu, možda posredstvom ugrađivanja u ribosom. Prema tome, ribosom je svojevrsni molekularni fosil koji u sebi sadržava dokaze koji podržavaju stereokemijsku teoriju.

4.2. KOEVOLUCIJSKA TEORIJA

Koevolucijska teorija, uzimajući u obzir vjerojatnu kronologiju pojavljivanja aminokiselina u repertoaru koda, je zapravo ideja da je kod u početku raspolagao s malim brojem aminokiselina dostupnih iz okoliša iz kojih su one ostale derivirane metaboličkim putevima, nakon čega su te kompleksnije aminokiseline ukomponirane u kod i pridruženi su im kodoni slični onima koji kodiraju za njihove prekursore. Dakle, struktura koda bi prema ovoj teoriji odražavala parove prekursor – produkt; kod i metabolički putevi su koevoluirali (Wong, 1975.).

Pretpostavljeni prekursor – produkt parovi su:

Glu → Gln	Asp → Lys	Ser → Trp
Glu → Pro	Gln → His	Ser → Cys
Glu → Arg	Thr → Ile	Val → Leu
Asp → Asn	Thr → Met	Phe → Tyr
Asp → Thr		

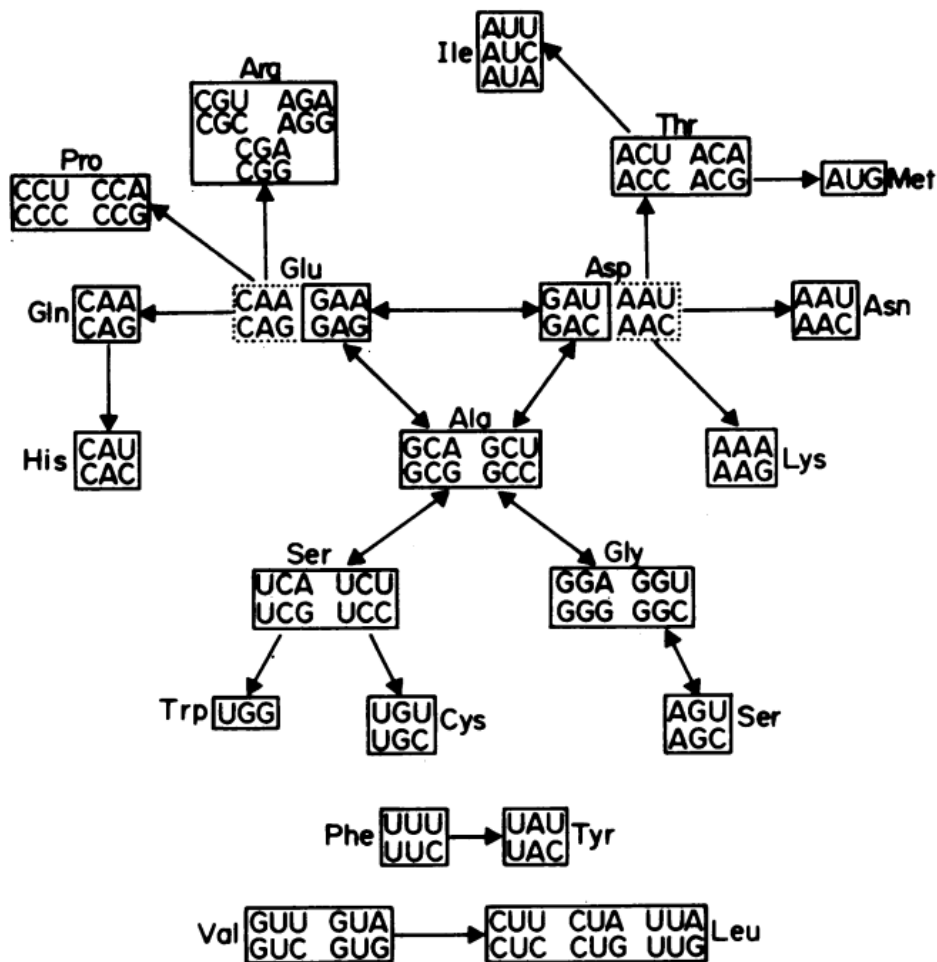
Nisu svi navedeni parovi potpuno definitivni. Na primjer, lizin može biti sintetiziran i od aspartata putem diaminopimelata i od glutamata putem alfa-aminoadipata (Wong, 2007.). S obzirom da prokarioti koriste diaminopimelatski put, aspartat se čini vjerojatnijim prebiotskim prekursorom lizina. Također, biosinteza metionina danas većinom počinje od aspartata, no put kojim se može dobiti od treonina sadrži manje metaboličkih koraka, a ako se homoserin uzme kao primitivni oblik treonina (Weber i Miller, 1981.), taj put ima još manje metaboličkih koraka.

Ako se sagleda tablica SGC s prethodno navedenim odnosima na umu, lako je primjetiti da mnogi od prekursor – produkt parova zauzimaju slične kodone (koji se razlikuju samo u jednom nukleotidu). Ako se doda ovoj teoriji još jedna pretpostavka – da su kodoni CAA i CAG pripadali glutamatu prije ugradnje glutamina u SGC, a AAU i AAC aspartatu prije ugradnje asparagina, moguće je sagraditi koevolucijsku mapu SGC i aminokiselina kao na Slici 4.

Neke od ovih odnosa moguće je objasniti i teorijom minimizacije greške; metabolički povezane aminokiseline često su kemijski slične te je njihova međusobna zamjena manje nepovoljna po fitnes od zamjene nekom jako različitom aminokiselinom. Međutim, mnogi prekursor – produkt parovi su kemijski izrazito različiti (Ser – Trp, Thr – Ile, Gln – His, Ile – Met, Asn – Lys, Cys – Trp), te se sličnost njihovih kodona ne može objasniti minimizacijom greške, s obzirom da njihova zamjena ne pruža značajne prednosti po fitnes u odnosu na nasumičnu zamjenu (Wong, 1975). Ovo također implicira i da kod, nakon prvotnog ustanovljenja, nije značajno adaptirao pod selekcijskim pritiscima (barem ne pod pritiskom minimizacije greške).

Međutim, ova teorija nije bez svojih kritičara. Pretpostavljeni prekursor – produkt parovi prema nekim istraživanjima (Amirnovin, 1997., Ronneberg, Freeland i Landweber, 2000.) nisu podržani suvremenim biokemijskim saznanjima, te na mjestima čak uzimaju energetske nepovoljne

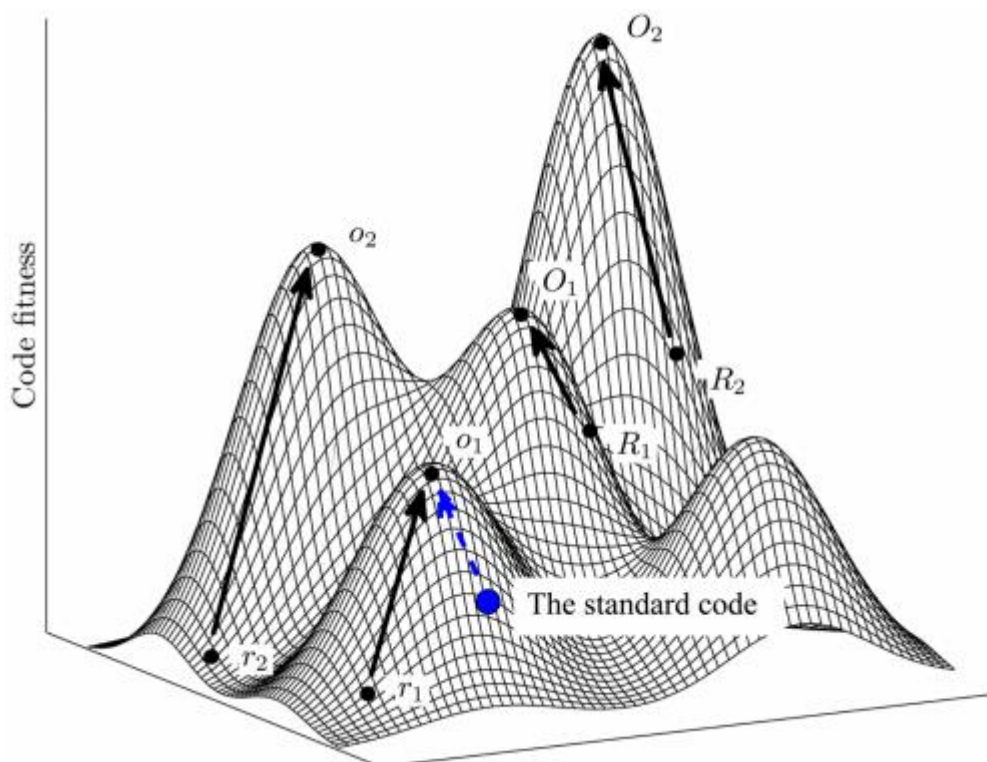
reakcije kao važne za ugrađivanje novih aminokiseline u kod, što možda pokazuje pretpostavku veće kompleksnosti metabolizma od onih kojima su raspolagale primitivne stanice. Također je kritizirana statistička metodologija rada u kojemu je teorija originalno izložena; prema jednom članku koji je koristio drugačiju metodologiju za određivanje vjerojatnosti da su ovakvi odnosi slučajni, čak 34% nasumičnih kodova pokazuje snažniju metaboličku korelaciju među kodonima od SGC, a ako se uzme u obzir i ideja da SGC ne sadrži 62, već 45 funkcionalnih kodona (nijedna poznata tRNA ne može razlikovati U i C na trećem mjestu kodona, pa su svi NNC i NNU kodoni translacijski ekvivalentni), statistička značajnost u korist ove teorije potpuno iščezava (Ronnenberg, Freeland i Landweber, 2000.). Unatoč tome, teorija je još uvijek relevantna i prema nekima (Wong, 2005.) spomenute kritike nisu potpuno utemeljene, pa se oko validnosti njezinih pretpostavki još uvijek odvija živa debata.



Slika 4. Koevolucijska mapa genetičkog koda. Puni okviri označavaju današnje kodone koji odgovaraju aminokiselinama napisanim pored, a iscrtkani okviri moguće kodone koji su nekada odgovarali aminokiselinama pored okvira. Strelice označavaju prekursor-produkt parove; jednostrane ireverzibilne reakcije a dvostrane reverzibilne reakcije. Svi parovi aminokiselina spojeni strelicama kodirani su kodonima koji se razlikuju u samo jednom nukleotidu. Preuzeto iz Wong, 1975.

4.3. TEORIJA MINIMIZACIJE GREŠKE

Posljednja od glavnih teorija koja će biti izložena u ovome radu jest teorija minimizacije greške, što je hipoteza da je prirodna selekcija pridružila biokemijski sličnim aminokiselinama slične kodone kako bi se umanjio efekt mutacija i pogrešaka u translaciji. Iako je teorija intuitivna, njezina provjera nije trivijalna zato što je kod računanja vjerojatnosti da SGC ima bolju minimizaciju greške od nasumičnog koda potrebno definirati funkciju gubitka, tj. Kvantificirati učinak mutacija i pogrešaka u translaciji na fitnes. Funkcija treba pridodati numeričku vrijednost 'cijene' zamjeni svake aminokiseline svakom od preostalih. S obzirom da ta cijena ovisi o kontekstu tj. Položaju i funkciji aminokiseline u proteinu, bilo koja definicija funkcije gubitka bit će nužno barem malo proizvoljna. Često su korištene metrike PR skala (kao na Slici 1.), indeks hidropatije i molekularni volumen (de Oliveira i sur., 2015.). Bitne su također i metode pretraživanja nasumičnih kodova kao i statistička metodologija obrade dobivenih podataka; osim prethodno spomenutog rezultata da je SGC 'jedan od milijun', korištenjem nešto drugačijih parametara kasnije su dobiveni i rezultati koji su još više obećavajući za teoriju minimizacije greške. Unatoč tome, od mogućih preko 1084 kodova mnogobrojni su više optimizirani, dapače, SGC se ne nalazi čak ni na lokalnom maksimumu funkcije optimizacije (Di Giulio i sur., 2001.).



Slika 5. Shema reljefa funkcije fitnesa kodova. SGC nalazi se otprilike na pola puta između optimalnosti nasumičnog koda i svog lokalnog vrhunca, što znači da ga je moguće daljnje optimizirati uz još jedan ili više evolucijskih koraka. Lokalni vrhunac također nije izvanredan te se njemu u blizini nalaze vrhunci većega fitnesa. Preuzeto iz (Novozhilov i sur., 2007.).

Postoji nekoliko problema s teorijom minimizacije greške, ili barem s učestalim metodama njezine provjere. Osim spomenute proizvoljnosti inherentne u izboru metrike sličnosti aminokiselina, često se kod sastavljanja funkcije gubitka pretpostavlja da je vjerojatnost svake supstitucije i pogreške u translaciji jednaka, dok je eksperimentalno utvrđeno da su tranzicije daleko češće od transverzija, kao i da su pogreške u translaciji prvoga i trećeg nukleotida u kodonu puno češće od pogrešaka u translaciji drugoga nukleotida u kodonu (Parker, 1989., Gilis i sur. 2001.). Također, najčešće se SGC uspoređuje s nasumičnim kodovima iste strukture - 64 kodona i 20 aminokiselina, a neki od radova pri pretraživanju prostora nasumičnih kodova čak koriste samo kodove strukturirane u iste sinonimne blokove kao SGC, isključujući čitav spektar mogućih različitih kodova (Salinas i sur., 2016.). Ovo možda i jesu donekle valjane metode za utvrđivanje robusnosti SGC, no ne govore puno o evolucijskom putu kojim je kod došao do današnjega oblika; robusnost može biti i nusprodukt porijekla SGC preko nekakvoga drugog puta. Kritičari teorije često napominju kako se robusnost može potpuno objasniti koevolucijskom teorijom (Massey, 2008.), no kao što je već objašnjeno, validnost te teorije uvelike ovisi o pretpostavljenom povijesnom redoslijedu pojave aminokiselina. Kako god bilo, valja barem pokušati kvantificirati razinu optimizacije SGC.

Najočitiji pristup bio bi usporediti SGC s potpuno optimiziranim hipotetskim kodom (Slika 6.) i prosječnom optimizacijom nasumičnih kodova, preko funkcije nazvane postotak minimizacije (minimization percentage, MP):

$$MP = (\phi_{\text{mean}} - \phi_{\text{stand}}) / (\phi_{\text{mean}} - \phi_{\text{min}})$$

pri čemu je ϕ_{mean} funkcija gubitka nasumičnih kodova, ϕ_{stand} funkcija gubitka SGC, a ϕ_{min} funkcija gubitka optimalnoga koda. Uzimajući PR skalu kao metriku sličnosti aminokiselina, za SGC dobiven je MP od 78% (Novozhilov i sur., 2007.). Optimalni kod dobiven je preraspoređivanjem kodona na način da je degeneracija i blok struktura koda održana, s obrazloženjem koje pretpostavlja da se takva struktura nekad u ranoj fazi evolucije SGC zamrzнула zbog radikalnoga delecijuskog učinka restrukturiranja koda.

Evolucijski putevi koji mogu dovesti do ove razine optimizacije istraživani su računalno, pretraživanjem reljefa funkcije fitnesa kodova. U jednome radu (Novozhilov i sur., 2007.), SGC je uspoređen s četiri skupa kodova: nasumični kodovi, skup nasumičnih kodova s većom razinom optimalnosti od SGC, te dva skupa kodova dobivenih optimizacijom prvih dvaju skupova. Dobiveno je da se razina optimalnosti prisutna kod SGC može postići, počevši od nasumičnog koda, u prosjeku preko samo 10-12 evolucijskih koraka (zamjena kodona), a ako se počne od optimalnijeg koda može se dobiti i puno veća razina optimiziranosti od SGC-ove. Dakle, SGC se nalazi otprilike na pola puta između optimalnosti nasumičnog koda i svog lokalnog vrhunca na reljefu fitnesa kodova. Uz to, taj lokalni vrhunac je osrednji u usporedbi s drugim obližnjim vrhuncima kao na Slici 5. S obzirom da

bi se SGC mogao daljnje optimizirati s još jednim ili više jednostavnih evolucijskih koraka, čini se da je ipak vjerojatno zapeo tu gdje danas jest barem dijelom zbog zamrznute slučajnosti.

UUU[E] Glu	UCU[S] Ser	UAU[Y] Tyr	UGU[F] Phe
UUC[E] Glu	UCC[S] Ser	UAC[Y] Tyr	UGC[F] Phe
UUA[D] Asp	UCA[H] His	UAA[] Ter	UGA[] Ter
UUG[D] Asp	UCG[H] His	UAG[] Ter	UGG[W] Trp
CUU[N] Asn	CCU[G] Gly	CAU[V] Val	CGU[C] Cys
CUC[N] Asn	CCC[G] Gly	CAC[V] Val	CGC[C] Cys
CUA[K] Lys	CCA[G] Gly	CAA[V] Val	CGA[L] Leu
CUG[K] Lys	CCG[G] Gly	CAG[V] Val	CGG[L] Leu
AUU[Q] Gln	ACU[A] Ala	AAU[T] Thr	AGU[L] Leu
AUC[Q] Gln	ACC[A] Ala	AAC[T] Thr	AGC[L] Leu
AUA[R] Arg	ACA[A] Ala	AAA[T] Thr	AGA[L] Leu
AUG[R] Arg	ACG[A] Ala	AAG[T] Thr	AGG[L] Leu
GUU[R] Arg	GCU[S] Ser	GAU[P] Pro	GGU[I] Ile
GUC[R] Arg	GCC[S] Ser	GAC[P] Pro	GGC[I] Ile
GUA[R] Arg	GCA[S] Ser	GAA[P] Pro	GGA[I] Ile
GUG[R] Arg	GCG[S] Ser	GAG[P] Pro	GGG[M] Met

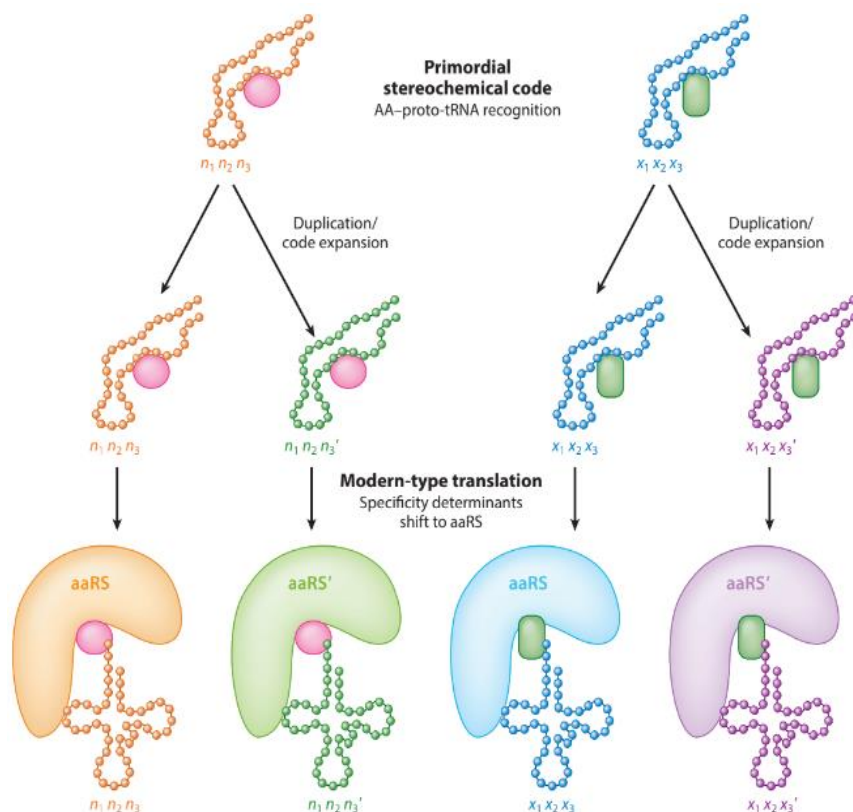
Slika 6. Optimizirani kod iste razine degeneracije i blok strukture kao SGC. Preuzeto iz Koonin i Novozhilov, 2009.

5. KOKOŠ ILI JAJE?

Iako je pomalo van raspona teme ovoga rada, ukratko ću izložiti problematiku koja se često razmatra odvojeno od evolucije genetičkoga koda, iako su teme iznimno usko povezane. Radi se, naime, o evoluciji translacijske mašinerije. Čini se paradoksalnom rana evolucija genetičkog koda: kako bi kod imao smisla, zahtjeva da postoji barem rudimentarna translacijska mašinerija koja je barem dijelom sastavljena od proteina; istovremeno, kako bi proteini postojano obavljali svoju funkciju kroz generacije, potrebno je da budu vjerno translatirani što zahtjeva već etabliran genetički kod. Što je nastalo prvo, kod ili translacija? Kokoš ili jaje?

Proteini koji su funkcionalni prevoditelji genetičkog koda su aaRS. Današnje aaRS podijeljene su u dvije paralogne klase, a svaka klasa sadrži različite aaRS za proizvodnju 10 različitih aminoacil-tRNA (Delarue, 1995.). aaRS klase 1 sadrže katalitički motiv Rossmanovoga nabora (Rossman fold), te time pripadaju široj obitelji proteina s Rossmanovim naborom (Wolf i sur., 2017.). Filogenetičkim istraživanjima te obitelji proteina pokazano je da su aaRS klase 1 relativno mlade pridošlice obitelji, što znači da je vjerna translacija u nekom obliku morala postojati daleko prije njihove evolucije i ustanovljenja kao prevoditelja genetičkog koda kako bi bila moguća evolucija proteinske obitelji s Rossmanovim naborom. Logičan zaključak izloženoga jest da je prije aaRS funkciju prevoditelja ispunjavala neka RNA.

Verzije ovakvog obrazloženja koriste stereokemijske teorije nešto drugačijega oblika od one objašnjenje u poglavlju 4.1. ovoga rada. Naime, nije nužno da stereokemijska specifičnost postoji baš između kodona/antikodona i aminokiseline; bilo koji dio tercijarne strukture proto-tRNA mogao je prepoznavati pripadajuće aminokiseline i vezati ih za aminoacil-vezujuće mjesto, kao na Slici 7.



Slika 7. Shema evolucije aminoacilirajućih ribozima - proto-tRNA, i preuzimanja funkcije prepoznavanja aminokiselina i aminoacilacije od strane aaRS. Preuzeto iz Koonin i Novozhilov, 2017.

Evolucijom od ovakvoga početka nije teško zamisliti kako ekspanzija repertoara aminokiselina kao posljedicu ima minimizaciju greške: duplikacijom proto-tRNA i modifikacijom afiniteta za novostečenu aminokiselinu prirodno će se dogoditi da sličnije aminokiseline vežu proto-tRNA slične tercijarne strukture i time sličnijega antikodona. Nakon što su aaRS preuzele funkciju prepoznavanja aminokiselina i aminoacilacije tRNA, svaki trag afiniteta tRNA za pripadajuće aminokiseline izgubio se kroz vrijeme, tako da pokusi s aptamerima teško da mogu polučiti definitivne rezultate ukoliko je opisani model točan (Kumar i Yarus, 2001.). Nada se možda nalazi u *in vitro* pokusima evolucije aminoacilirajućih ribozima, što su pokusi koji bi, ako bi bili uspješni, mogli dobrim dijelom rekapitulirati čitav proces evolucije genetičkog koda.

6. ZAKLJUČAK

Unatoč tome što se na pitanje porijekla genetičkog koda pokušava odgovoriti već pola stoljeća, ne čini se da je znanstvena zajednica prišla blizu definitivnom odgovoru. Ovaj problem čini se jedinstvenim po tome koliko je važan, zanimljiv i iznad svega temeljan za evolucijsku biologiju, ali s druge strane inherentno zahtjevan za rješavanje. Stereokemijska teorija nije nikada polučila uvjerljivim rezultatima, barem u obliku u kojemu je uobičajeno postavljena; teorija minimizacije greške razotkrila je važne karakteristike SGC, no ne odgovara na pitanja porijekla i mehanizma postanka tih karakteristika; koevolucijska teorija zahtjeva još mnogo empirijske provjere i integracije s modernim biokemijskim saznanjima kako bi se mogla pomaknuti od trenutne mrtve točke. Čini se da se budućnost rješavanja ovoga problema nalazi u integriranom pristupu koji u obzir uzima sve spomenute potencijalne evolucijske putove i selekcijske pritiske, što je model sa zastrašujuće puno parametara za provjeru, kao i evoluciju translacijske mašinerije, također izazovno područje zbog svoje univerzalne konzerviranosti.

7. SAŽETAK

Standardni genetički kod (SGC) je šifra prevođenja sekvence nukleotida u sekvencu aminokiselina na način da je svaki od mogućih tripleta (nizova od 3) nukleotida – kodon pridružen nekoj od 20 bioloških aminokiselina. SGC je gotovo univerzalan u živom svijetu, zbog čega znamo da je kao takav postojao i u posljednjem univerzalnom zajedničkom pretku (LUCA-i). Unatoč tome što je SGC najkonzerviranija informacija u biologiji, pitanje njegovoga porijekla do danas je nerazriješeno. U ovome radu izložene su vodeće teorije o evoluciji SGC: stereokemijska teorija koja postulira da je stereokemijski afinitet između kodona ili antikodona i pripadajuće aminokiseline uzrok njihovome pridruživanju; koevolucijska teorija koja kaže da je kod prvotno raspolagao s malim brojem aminokiselina prisutnih u okolišu, nakon čega su ostale kasnije dobivene metaboličkim putevima iz potonjih i zatim uključene u kod pri čemu su im pridruženi kodoni srodni onima njihovih prekursora; teorija minimizacije greške koja tvrdi da je otpornost stanice na mutacije i pogreške u translaciji bila predvodeći selekcijski pritisak koji je gurao evoluciju SGC. Svaki od spomenutih pristupa ima u obliku u kojemu su najčešće izloženi mane koje se čine nepremostivima te se budućnost istraživanja porijekla SGC vjerojatno nalazi u integriranom eksperimentalnom pristupu koji u obzir uzima sve od spomenutih faktora, a i evoluciju aaRS i ostatka translacijske mašinerije, kao relevantne u ranoj evoluciji i porijeklu SGC.

8. SUMMARY

The standard genetic code (SGC) is a code by which sequences of nucleotides are translated into sequences of amino acids such that each nucleotide triplet (sequence of three) called a codon is assigned to any one of 20 biological amino acids. The SGC is nearly universal in all extant life forms, which is why we can safely presume it had been established as such at least as far back as the last universal common ancestor (LUCA). Although the SGC is perhaps the greatest information constant in all of biology, the question of its origin remains unanswered. In this paper, main theories of its evolution and origin are demonstrated: the stereochemical theory, which hypothesizes that a stereochemical affinity was the driving force of adjoining codons to their cognate amino acids; the coevolution theory which posits that the code originally had fewer amino acids at its disposal, and the rest of the 20 were derived from this original set via assorted metabolic pathways, after which codons that previously had belonged to their precursors were assigned to them; the error minimization theory which says that the pressure to resist the deleterious effects of mutations and translation errors was the main selective pressure to drive SGC evolution. Each of the described theories or the approaches by which they are tested have their flaws, so the future of research on the origin and evolution of the SGC appears to be in an integrated approach which takes into account all of the above, as well as the evolution of the translation apparatus, as relevant in the early evolution of the SGC.

9. LITERATURA

- Amirnovin R. 1997. An Analysis of the Metabolic Theory of the Origin of the Genetic Code. *Journal of Molecular Evolution*. 44(5):473–476.
- Ardell D., Sella G. 2002. No accident: Genetic codes freeze in error-correcting patterns of the standard genetic code. *Philosophical Transactions of The Royal Society B Biological Sciences* 357(1427):1625-42.
- Bada JL. 2013. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. *Chemical Society Reviews*. 42 (5): 2186–96.
- Brooks DJ., Fresco JR., Lesk AM., Singh M. 2002. Evolution of Amino Acid Frequencies in Proteins Over Deep Time: Inferred Order of Introduction of Amino Acids into the Genetic Code. *Molecular Biology and Evolution*, 19(10):1645–1655.
- Crick FH. 1968. The origin of the genetic code. *J. Mol. Biol.* 38:367–79.
- de Oliveira LL., de Oliveira PS., Tinos R. 2015. A multiobjective approach to the genetic code adaptability problem. *BMC Bioinform.* 16:52.
- Delarue, M. 1995. Aminoacyl-tRNA synthetases. *Structural Biology*. 5: 48–55.
- Di Giulio M., Medugno M. 2001. The level and landscape of optimization in the origin of the genetic code. *J. Mol. Evol.* 52:372–82.
- Freeland SJ., Hurst LD. The genetic code is one in a million. *J Mol Evol.* 47(3):238-48.
- Gamow, G. 1954. Possible relation between deoxyribonucleic acid and protein structures. *Nature* 173, 318.
- Gilis D., Massar S., Cerf NJ., Rooman M. 2001. Optimality of the genetic code with respect to protein stability and amino-acid frequencies. *Genome Biol.* 2(11): research0049.1–research0049.12.
- Ibba M. i sur. 2010. tRNAs: cellular barcodes for amino acids. *FEBS Lett.* 584(2): 387–395.
- Johnson DBF., Wang L. 2010. Imprints of the genetic code in the ribosome. *Proc Natl Acad Sci U S A.* 107(18): 8298–8303.
- Kasting JF. 1993. Earth's Early Atmosphere. *Science*. 259 (5097): 920–926.
- Keeling PJ. 2016. Genomics: Evolution of the Genetic Code. *Current Biology* 26(18).
- Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet.* 2:49–58.
- Knight RD., Landweber LF. 1998. Rhyme or reason: RNA-arginine interactions and the genetic code. *Chem Biol.* 5(9):215–220.
- Knight RD., Landweber LF., Yarus M. 2003. Tests of a Stereochemical Genetic Code. *Translation Mechanism*. Kluwer Academic/Plenum Publishers; New York: 115–128.
- Kobayashi K., Tsuchiya M., Oshima T., Yanagawa H. 1990. Abiotic synthesis of amino acids and imidazole by proton irradiation of simulated primitive earth atmospheres. *Origins of Life and Evolution of the Biosphere.* 20(2):99–109.

- Koonin EV., Novozhilov AS. 2009. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life*. 61(2): 99–111.
- Kumar RK., Yarus M. 2001. RNA-catalyzed amino acid activation. *Biochemistry*. 40:6998–7004.
- Levin BR., Bergstrom CT. 2000. Bacteria are different: Observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc Natl Acad Sci U S A*. 97(13): 6981–6985
- Martin WF., Nelson-Sathi S. 2016. The physiology and habitat of the last universal common ancestor. *Nature Microbiology* volume 1, Article number: 16116.
- Massey SE., 2008. A neutral origin for error minimization in the genetic code. *J. Mol. Evol.* 67:510–16.
- Miller SL. 1953. Production of Amino Acids Under Possible Primitive Earth Conditions. *Science*. 117 (3046): 528–9.
- Miller SL., Urey, HC. 1959. Organic Compound Synthesis on the Primitive Earth. *Science*. 130 (3370): 245–51.
- Novozhilov AS., Wolf YI., Koonin EV. 2007. Evolution of the genetic code: partial optimization of a random code for robustness to translation error in a rugged fitness landscape. *Biol Direct*. 2: 24.
- Oparin AI. 1938. *The Origin of Life*
- Parker J. 1989. Errors and alternatives in reading the universal genetic code. *Microbiol Rev*. 53:273–298.
- Proteins. *J Mol Evol* 60:128–139
- Ronneberg TA., Landweber LF., Freeland SJ. 2000. Testing a biosynthetic theory of the genetic code: Fact or artifact? *Proceedings of the National Academy of Sciences*. 97(25):13690–13695.
- Salinas DG., Gallardo MO., Osorio MI. 2016. Local conditions for global stability in the space of codons of the genetic code. *Biosystems* 150:73–77.
- Sengupta S., Higgs PG. 2005. A Unified Model of Codon Reassignment in Alternative Genetic Codes. *Genetics*. 170(2): 831–840
- Sengupta S., Higgs PG. 2015. Pathways of Genetic Code Evolution in Ancient and Modern Organisms. *J Mol Evol*. 80(5-6):229-43
- Sneath PH. 1966. Relations between chemical structure and biological activity in peptides. *Journal of Theoretical Biology*. 12(2):157-95.
- Swire J., Judson OP., Burt A. 2005. Mitochondrial Genetic Codes Evolve to Match Amino Acid Requirements of
- Taylor FJ., Coates D. 1989. The code within the codons. *Biosystems*. 22(3):177-87
- Trifonov EN. 2000. Consensus temporal order of amino acids and evolution of the triplet code. *Gene*. 261(1):139–151
- Weber AL., Miller SL. 1981. Reasons for the occurrence of the twenty coded protein amino acids. *J. Mol. Evol*. 17:273-284.

Wolf YI., Aravind L., Grishin NV., Koonin E. 2017. Evolution of Aminoacyl-tRNA Synthetases—Analysis of Unique Domain Architectures and Phylogenetic Trees Reveals Complex History of Horizontal Gene Transfer Events. *Genome*. 9:689–710.

Wong JT. 2005. Coevolution theory of the genetic code at age thirty. *Bioessays*. 27(4):416-25.

Wong JTF. 1975. A Co-Evolution Theory of the Genetic Code. *Proceedings of the National Academy of Sciences*. 72(5):1909–1912

Wong JTF. 2007. Question 6: Coevolution Theory of the Genetic Code: A Proven Theory. *Origins of Life and Evolution of Biospheres*. 37(4):403–408.

https://en.wikipedia.org/wiki/Genetic_code#cite_note-isbn0-674-05075-4-66